UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001
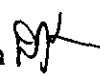
March 8, 2000   **MASTER FILE**

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES R-26

MEMORANDUM FOR Maureen Lynch
                Assistant Division Chief, Coverage Measurement Processing
                Decennial Statistical Studies Division

From: Donna Kostanich
                Assistant Division Chief, Sampling and Estimation
                Decennial Statistical Studies Division

Prepared by: Ryan Cromar
                Sample Design Team
                Decennial Statistical Studies Division

Subject: Accuracy and Coverage Evaluation: Large Block Cluster Subsampling
                Parameter File Specifications

I.   INTRODUCTION

This memorandum provides specifications for the large block cluster subsampling
parameter file for the Census 2000 Accuracy and Coverage Evaluation (A.C.E.) survey.
As the final stage of the A.C.E. design, large block cluster subsampling involves selecting
a portion of a block cluster that has 80 or more A.C.E. housing units to be in the A.C.E.
interview sample. This will be accomplished by forming segments of adjacent housing
units within the block cluster and selecting a subsample of segments. The objective of
large block cluster subsampling is to meet the target A.C.E. interviewing sample sizes
using the most up-to-date A.C.E. housing unit counts available at the time of
subsampling. The creation of the parameter file is the first step of large block cluster
subsampling. The parameter file is an input for the large block cluster subsampling
process documented in reference 1. The large block cluster subsampling parameter file
specification is similar to the specifications prepared for the 1998 Census 2000 Dress
Rehearsal and documented in reference 2.

Earlier stages of the A.C.E. sample design include the selection of A.C.E. block clusters
for the listing sample (see reference 3), the A.C.E. reduction (see reference 4), and the
subsampling of small block clusters (see reference 5). After the listing sample selection,
the independent list is created and the results keyed and verified. Based on the results of

this listing, the A.C.E. sample reduction and small block cluster subsampling are done. Subsequently, the housing unit matching and follow-up are done, and the preliminary enhanced list is created and sent to large block cluster subsampling. The preliminary enhanced list is both the input and output file for large block cluster subsampling. The output preliminary enhanced list is updated with the subsampling results, and is referred to as the subsampled preliminary enhanced list. The enhanced List is created by extracting only the housing units designated for interview following large block cluster subsampling from the subsampled preliminary enhanced list (see reference 6).

This memorandum is organized into the following sections:
- Assumptions
- Definitions
- Input
- Process
- Output
- Verification
- References

These specifications should be used to flowchart the process, to generate further discussion on requirements, to identify and finalize the record layouts of input and output files, and to write computer software to implement the methodology. During and after a testing phase, it is possible that changes to the specifications will be necessary.

If there are any questions or comments, please contact Ryan Cromar (301-457-1636), James Farber (301-457-4282), or Deborah Fenstermaker (301-457-4195) of the Decennial Statistical Studies Division (DSSD).

II.   ASSUMPTIONS

The assumptions required to create the large block cluster subsampling parameter file are:

A.   The creation of the large block cluster subsampling parameter file is not affected by the results of housing unit matching and follow-up. Thus the creation of this parameter file can be done using independent list housing unit counts at any time after small block cluster subsampling even if matching and follow-up are not completed. It is possible that the housing unit counts determined in this specification will differ due to housing unit follow-up.

B.   Block clusters eligible for large block cluster subsampling include those that were selected in the A.C.E. reduction and remain in the sample following small block cluster subsampling. All other block clusters are not eligible for large block cluster subsampling.

C. Large block cluster subsampling is done on a flow basis over a span of several days. Therefore, the large block cluster subsampling parameter file includes fields for daily statistics that will be filled during the subsampling process and will be used to track the day-to-day results of large block cluster subsampling.

D. The large block cluster subsampling parameter file will not be revised to account for relisted block clusters. Relisted block clusters will receive the original subsampling parameters computed from the independent list before relisting. Block clusters which require relisting will not be identified nor will the relisting be done by the time the large block cluster subsampling parameters are calculated.

E. The A.C.E. housing units on the independent list are keyed and valid.

F. The A.C.E. housing units that have a Unit Status of Future Construction are excluded from the take-every calculation. Including such units could cause the A.C.E. interview sample size to be lower than expected, which would increase the variance of the A.C.E. population estimates. Excluding these units is a conservative approach to ensure that target sample sizes are achieved.

G. All decimal numbers are rounded to six digits at the time of creation using the standard rounding procedure except when noted otherwise. Decimal numbers with a seventh decimal place of five or more are rounded up in the sixth decimal place. Those with four or less in the seventh decimal place are rounded down in the sixth decimal place.

H. Note that medium and small block clusters are eligible for large block cluster subsampling since the decision to subsample is based only on the number of A.C.E. housing units in the block cluster.

III. DEFINITIONS

A. American Indian Reservation (AIR) Block Cluster

A block cluster with three or more housing units based on information available at the time block clusters were formed that is at least partially in an AIR. See Sampling Strata.

B. Block Cluster

A geographically contiguous group of Census 2000 collection blocks (see reference 7).

3

C.   Listing Sample

The initial sampling stage of A.C.E. in which block clusters are selected for independent listing (see reference 3).

D.   A.C.E. Independent List

List of all housing units in A.C.E. listing sample block clusters. The independent list is created independently of the Decennial Master Address File, the address list used for the census.

E.   Keyed and Valid Housing Units

Housing units with any of the following Unit Status codes on the independent list:

        1 = Occupied or vacant and intended for occupancy
        2 = Under construction
        3 = Future construction
        4 = Unfit for habitation
        5 = Boarded up
        6 = Storage of household goods
        7 = Vacant mobile home site
        8 = Other

All of these units are included because it is possible that the unit status may change between listing and interviewing. Group quarters are not listed in A.C.E. Note that A.C.E. Future construction housing units (Unit Status = 3) are excluded from the take-every calculation as explained in section II above.

F.   Large Block Cluster

See Sampling Strata.

G.   Medium Block Cluster

See Sampling Strata.

H.   A.C.E. Housing Unit

A housing unit on the preliminary enhanced list that is keyed and valid and has one of the following After Follow-up Match Codes: M, MU, UI, or CI.

I. A.C.E. Reduction

The process of reducing the A.C.E. listing sample from the Integrated Coverage Measurement (ICM) sample to the A.C.E. interview sample. In the A.C.E. reduction, the listing sample block clusters are subsampled, and the selected block clusters continue to small block cluster subsampling (see Reference 4).

J. A.C.E. Reduction Strata

A partition (mutually exclusive and exhaustive set) of all block clusters in a state into groups according to certain characteristics. See Attachment A for a list of the A.C.E. Reduction Strata and see reference 4 for more information on how A.C.E. reduction strata are defined.

K. Sampling Strata

A partition of all block clusters within a state into groups according to the number of housing units estimated in each cluster at the time of block clustering (see reference 7). Block Clusters are assigned to sampling strata prior to listing sample selection. The sampling strata are:

1	=	small block clusters with 0 - 2 estimated housing units
2	=	medium non-AIR block clusters with 3 - 79 estimated housing units
3	=	large non-AIR block clusters with ≥ 80 estimated housing units
4	=	medium and large AIR block clusters with ≥ 3 estimated housing units

L. Small Block Cluster

See Sampling Strata.

M. State

The 50 United States plus the District of Columbia and Puerto Rico.

IV.   INPUT FILES

The inputs for the subsampling process are the following:

A.   Large Block Cluster Subsampling Input File

Description: This file contains the target housing unit sample size for each
A.C.E. reduction stratum within each state.
Level: A.C.E. reduction stratum
Scope: One record per A.C.E. reduction stratum within each state
Layout: See Attachment B

B.   Cluster Status File

Description: This file has one record for each of the 29,695 block clusters
selected in the A.C.E. listing sample. It is updated with
information from other processing stages. For large block cluster
subsampling, this file is used to determine the sampling
parameters.
Level: A.C.E. Block Cluster
Scope: All block clusters selected for the A.C.E. listing sample

C.   A.C.E. Sample Design File (Version 3)

Description: This file reflects the previous A.C.E. sampling operations: listing
sample selection, A.C.E. reduction, and small block cluster
subsampling.
Level: Block Cluster
Scope: One record for each block cluster in the A.C.E. listing sample
File Layout: See Attachment C

V.   PROCESS

Using results from the independent list, subsampling parameters for each A.C.E.
reduction stratum within each state are calculated prior to subsampling any block cluster.
The sampling parameters to calculate are the take-every, the target number of segments in
a block cluster, and the random start. All block clusters in a common A.C.E. reduction
stratum within each state have the same parameters.

A.   Determine the take-every by computing the number of keyed and valid housing
units on the independent list in all block clusters with 80 or more A.C.E. housing
units and dividing it by the target sample size from the block clusters with 80 or

more housing units, but excluding the number of future construction housing units from the calculation. Block clusters from the small sampling stratum with fewer than 10 housing units on the independent list are not included when calculating housing unit totals because these housing units are not included in the target sample size.

Use the following steps to determine the take-every for each A.C.E. reduction stratum with each state:

1.   Obtain four housing unit counts for each A.C.E. reduction stratum within each state from the cluster status file. Exclude Future Construction housing units, those with Unit Status = 3, from all of these counts.

     a.   The number of housing units in all block clusters from the independent list, NILHUT.

     b.   The number of housing units in block clusters with 80 or more housing units from the independent list, NILHUL.

     c.   The number of housing units in small block clusters (sampling stratum = 1) with fewer than ten housing units from the independent list, NILHUS.

     d.   The number of housing units in block clusters with fewer than 80 housing units from the independent list excluding block clusters from step c above, NILHUM.

     e.   Calculate Z to check the counts above. This is calculated by subtracting NILHUL, NILHUS and NILHUM from NILHUT.

     $$Z = NILHUT - (NILHUL + NILHUS + NILHUM)$$

     If the counts are correct, Z will be equal to zero. Resolve the cases where Z is not equal to zero.

2.   Obtain the target number of sample housing units, T, for the A.C.E. reduction stratum within state from the large block cluster subsampling input file ACE2000_LBINPUT.FIN.

7

3. Calculate the take-every, TELB:

$$TELB = \frac{NILHUL}{T - NILHUM}$$

- If TELB < 1.012660[1], then set TELB = 1.000000.

- If T - NILHUM = 0, then contact the Sample Design Team of the DSSD.

4. Round the take-every to six decimal places.

B. Calculate the integer number of segments to be formed in a block cluster, NSEG, using the formulas below. These formulas are based on the TELB to ensure at least one segment is selected from each block cluster eligible for subsampling. In addition, create a variable called FORMULA, that codes which formula was used for calculating the number of segments.

- If TELB ≥ 2, then
    NSEG = TELB (If not an integer, round up to the next integer.)
    FORMULA = 1

- If 1 < TELB < 2, then

    $$NSEG = \frac{1}{1 - \frac{1}{TELB}}$$ (If not an integer, round up to the next integer.)

    FORMULA = 2

- If TELB = 1, then
    NSEG = 1
    FORMULA = 3

---

[1]This value for TELB was determined to prevent any block clusters from having over 80 segments.

8

C.   Calculate the random start, RS, by generating a random number, RN, between zero and one, rounding it to six decimal places, and multiplying it by the TELB. Round the resulting random start to six decimal places. Calculate a new random start for each A.C.E. reduction stratum within each state.

- $RS = RN \times TELB$, where $0 < RN \leq 1$

- If TELB = 1.000000, then set RS = 1.000000.

D.   Starting with the large block cluster subsampling input file as a basis, create a large block cluster sampling parameter file by appending the variables created in this section plus two other variables described in step 1 below. This large block cluster subsampling parameter file will be a daily input to the large block cluster subsampling. Additional variables will be appended during future steps in the large block cluster subsampling process. This file has one record for each A.C.E. reduction stratum within each state.

1.   Create two variables, current daily start, DS, and cumulative cluster count, CCC, which are needed for implementing the subsampling over several days. Initialize these variables by setting DS equal to the random start, RS, and assigning a value of zero to CCC for each A.C.E. reduction stratum within each state.

2.   Update the following variables on the large block cluster subsampling parameter file using the layout in Attachment D:

> Number of housing units in block clusters with 80 or more housing units on the independent list, NILHUL
> Number of housing units in block clusters with 0-79 housing units on the independent list (except smalls with 0-9), NILHUM
> Number of housing units in all block clusters on the independent list, NILHUT
> Number of housing units in small block clusters with 0-9 housing units on the independent list, NILHUS
> Take-every for the segment subsampling, TELB
> Number of segments in a block cluster, NSEG
> Flag for formula used for calculating NSEG, FORMULA
> Random Number between 0 and 1, RN
> Random Start for the segment subsampling, RS
> Current daily start for the segment subsampling, DS
> Cumulative Cluster Number, CCC

Round RS, DS and TELB to six decimal places. The other variables are integers.

E.   As soon as the parameter file is created, provide it to the Sample Design Team in the DSSD for review and approval prior to large block cluster subsampling.

9

VI.   OUTPUT FILE

The output requested by the Sample Design Team in the DSSD from the Process section is the following:

A.   The Large Block Cluster Subsampling Parameter File

Description: This file contains information needed for selecting the systematic subsample on a flow basis. The file will be produced after the sampling parameters are calculated. The final version will be created when the large block cluster subsampling process is complete.

Level: A.C.E. Reduction Stratum

Scope: One record per A.C.E. reduction stratum within each state

File Layout: See Attachment D

VII.   VERIFICATION

The following information should be provided for verification:

•   Large Block Cluster Subsampling Parameter File

•   Cluster Status File

Provide the sampling parameter file. See section VI above for information about this file. Access to the cluster status file is also required for verification.

VIII. REFERENCES

1   DSSD Census 2000 Procedures and Operations Memorandum Series R-27, "Census 2000 Accuracy and Coverage Evaluation: Large Block Cluster Subsampling Specifications," March 8, 2000.

2   DSSD Census 2000 Dress Rehearsal Memorandum Series A-9, "Census 2000 Dress Rehearsal ICM Sampling: Large Block Subsampling Specification," April 15, 1998.

3   DSSD Census 2000 Procedures and Operations Memorandum Series R-3, "Accuracy and Coverage Evaluation (ACE) Survey: Block Cluster Sample Selection Specification," March 29, 1999.

4   DSSD Census 2000 Procedures and Operations Memorandum Series R-, "Accuracy and Coverage Evaluation Survey: Reduction Specification," January 10, 2000, DRAFT.

5   DSSD Census 2000 Procedures and Operations Memorandum Series R-24, "Accuracy and Coverage Evaluation Survey: Small Block Cluster Subsampling," February 1, 1999.

6   DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-HU-08, "Creation of the Census 2000 Accuracy and Coverage Evaluation (A.C.E.) Enhanced List for Person Phase Interviewing," June 21, 1999, DRAFT.

7   DSSD Census 2000 Procedures and Operations Memorandum Series R-8, "Census 2000 Specifications for Block Cluster Formation-Reissue," May 3, 1999.

cc: DSSD Census 2000 Procedures and Operations Memorandum Series Distribution List
        A.C.E. Implementation Team Leaders
        Statistical Design Team Leaders
        Sample Design Team

A.C.E. Reduction Strata

Stratum Code[1]	Stratum Name
01	Minority
02	Non-minority Low Inconsistent
03	Non-minority Consistent
04	Non-minority High Inconsistent
05	Non-minority Inconsistent
06	Non-minority
07	Low Inconsistent
08	Consistent
09	High Inconsistent
10	Inconsistent
11	Minority Inconsistent
12	Minority Consistent
13	Full Collapse
14	Minority Low Inconsistent
15	Minority High Inconsistent
16	Medium Stratum Jumpers
17	American Indian Reservations
18	Puerto Rico
19	Small Stratum Jumpers

[1] Only Strata 01, 02, 03, 04, 16, 17, 18, and 19 were actually used for the A.C.E. Reduction. When developing the computer specifications for the A.C.E. cluster reduction and large block cluster subsampling, the cluster reduction design had not been determined. Thus, to accommodate several potential reduction design plans, we specified 19 strata, but only used eight.

Large Block Cluster Subsampling Input File Layout

Variable Description	Name	Pos
State	ST	1-2
A.C.E. reduction stratum	ARST	4-5
Target housing unit sample size	T	7-14

Sample Design File

The Sample Design File contains one record per block cluster selected during the listing sample selection. If the block cluster falls out of sample during the second step of the listing sample, the A.C.E. reduction, small block cluster subsampling, or the A.C.E. reduction, the remaining variables will be left blank. The initial version of the file, which will be created following the initial block cluster selection, is called SDF.US1. For each subsequent update to the file, the version number will increase by one (i.e. SDF.US2, SDF.US3). The layout for the Sample Design File is as follows:

Variable Description	Name	Places	Source
Census Region	REGION	1	UN
Census Division	DIV	2	UN
State code	STATE	3-4	UN
County code	COUNTY	5-7	UN
Local census office	LCO	8-11	CS
Interim Tract (Pseudo Tract)	ITRACT	12-17	BC
Current Sample Indicator	CSI	19	UO
A.C.E. block cluster number	CLUST	21-25	CS
Check Digit	DIGIT	26	CS
Geography block cluster number	GCLUST	28-32	BC
List/Enumerate Indicator	LEIND	33	BC
Type of Enumeration Area Recode	TEACR	34	CS
Type of Enumeration Area group	TEAG	36	BC
Number of HUs used for sample design	NHU	37-41	BC
Number of MAF HUs	NHUM	43-47	BC
Number of 1990 HUs	NHU90	49-53	BC
Sampling Stratum	SS	55	UN

      1 = Small
      2 = Medium
      3 = Large
      4 = American Indian Reservation

American Indian Country Indicator	AICIND	56	BC

      0 = No American Indian Country
      1 = American Indian Reservation/trust land
      2 = Tribal Jurisdiction Area/
          Alaska Native Village Statistical Area/
          Tribal Designated Statistical Area

Demographic/Tenure Group code	DTCODE	57-58	UN
Demographic/Tenure Group label	DTLABEL	59-60	UN
Estimated Urbanicity of block cluster	ECLUSURB	62	UN

      1 = Urban Area with population ≥250,000
      2 = Other Urban Area
      3 = Non-Urban Area

Size Category	SIZCAT	63	UN

      1=Small (0-2 hus)
      2=Medium (3-79 hus)
      3=Large (80+ hus)

Additional space		64-91	

Variable Description	Name	Places	Source
First step index number	INDEX1	92-99	CS
Listing sample selection indicator	BC1	101	CS
1 = Selected			
Random Start for listing sample selection	RS1	103-113	UN
Take-every for listing sample selection	TE1	115-125	UN
Second step listing sample selection indicator	BC2	127	CS
0 = Not Selected			
1 = Selected			
Random Start for the second step of the listing sampling	RS2	129-139	CS
Take-every for the second step of the listing sampling	TE2	141-151	CS
Unbiased weight after block cluster sampling	WEIGHTBC	153-164	CS
Additional space		165-175	

Variable Description	Name	Places	Source
Preliminary Number of HUs on the Independent List	NHUILP	176-180	AR
Number of Housing Units On the January 2000 DMAF	NHUDMAF	182-186	AR
Demographic Code	DEMCODE	188	AR
1 = Minority			
2 = Non-Minority			
3 = Puerto-Rico			
Consistency Code	CONCODE	189	AR
1 = Low Inconsistent (IL significantly smaller than DMAF)			
2 = Consistent			
3 = High Inconsistent ((IL significantly larger than DMAF)			
A.C.E. Reduction Stratum	ARST	190-191	AR
A.C.E. Reduction Indicator	ACERED	193	AR
0 = Not Selected			
1 = Selected			
Random Start for A.C.E. Reduction	RSAR	195-205	AR
Take-every for A.C.E. Reduction	TEAR	207-217	AR
Unbiased weight after A.C.E. reduction	WEIGHTAR	219-230	AR
Collapsing Flag	COLFLAG	232	AR
A.C.E. Reduction Index Number	INDEXR	234-241	AR
Number of Housing Units On the December 1999 DMAF (Initial)	NHUDMAFI	243-247	AR
Additional space		248-300	

Variable Description	Name	Places	Source
Number of HUs on the Independent List	NHUIL	301-305	SB
Small Block Cluster Subsampling Stratum	SBCSS	306-307	SB
Small Block Subsampling Indicator	SB	308	SB
0 = Not Selected			
1 = Selected			
Random Start for Small Block subsampling	RSSB	310-320	SB
Initial take-every for Small Block subsampling	ITESB	322-332	SB
Unbiased weight for A.C.E. cluster	WEIGHTC	334-345	SB
Larger of the DMAF and IL HU count	LARGERHU	347-351	SB
Final take-every for Small Block subsampling	FTESB	352-362	SB
Additional space		363-370	

2

Variable Description	Name	Places	Source
Relisted Block Cluster Flag	RELIST	371	LB
0 = Not Relisted, 1 = Relisted			
Number of total hus in block cluster	NHUEL	373-377	LB
Number of A.C.E. hus in cluster	NHUELA	379-383	LB
Number of supplemental hus in cluster	NHUELN	385-389	LB
Large Block Cluster EL subsampling code	ELLBSUB	391	LB
1 = NHUELI< 80 hus, 2 = NHUELI ≥ 80 hus			
Random Start for Large Block subsampling	RSLB	393-403	LB
Take-every for Large Block subsampling	TELB	405-415	LB
Number of segments in block cluster	NSEG	417-418	LB
Number of segments selected in block cluster	NSEGSAM	420-421	LB
Day of Arrival	DAY	423-424	LB
Final Cluster Order Number	CON	431-434	LB
Number of total hus for interview in block cluster	NINT	436-440	LB
Unbiased weight for P-sample HUs	WEIGHTP	442-453	LB
Number of Assignments in block cluster	NA	455-456	LB
Final Sampling Strata	FSS	458-464	LB
Additional space		465-490	

Source Codes

- AR: A.C.E. Reduction
- BC: Block Clustering
- CS: Block Cluster Sampling
- LB: Large Block Subsampling
- SB: Small Block Subsampling
- UN: Universe File Creation
- UO: Updated for each operation

Large Block Cluster Subsampling Parameter File Layout

Variable Description	Name	Pos
State	ST	1-2
A.C.E. reduction stratum	ARST	4-5
Target housing unit sample size	T	7-14
Number of housing units in block clusters with 80 or more housing units on the independent list	NILHUL	16-21
Number of housing units in block clusters with 0-79 housing units on the independent list (except smalls with 0-9)	NILHUM	23-28
Number of housing units in all block clusters on the independent list	NILHUT	30-35
Number of housing units in small block clusters with 0-9 housing units on the independent list	NILHUS	37-42
Take-every for the segment subsampling	TELB	44-54
Number of segments in a block cluster	NSEG	56-57
Flag for formula used for calculating NSEG	FORMULA	59
Random Number between 0 and 1	RN	61-72
Random Start for the segment subsampling	RS	74-84
Current Daily Start	DS	86-96
Cumulative Cluster Count	CCC	98-100
Daily Start for Day 1	DS1	102-112
Daily Start for Day 2	DS2	114-124
.	.	.
.	.	.
.	.	.
Daily Start for Day 20[1]	DS20	

I

---

[1] The number of days for sampling may be over or under 20. If this is the case, appropriate modifications will be made.